

技術者のための StatWorks/V5 機械学習編の 機能および有効な使い方

株式会社 日本科学技術研修所

数理事業部

犬伏秀生

本発表の説明項目

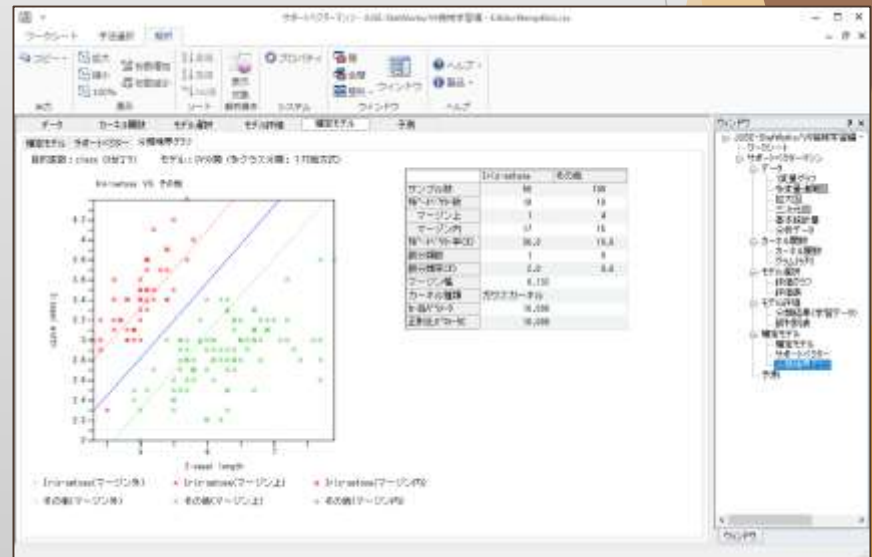
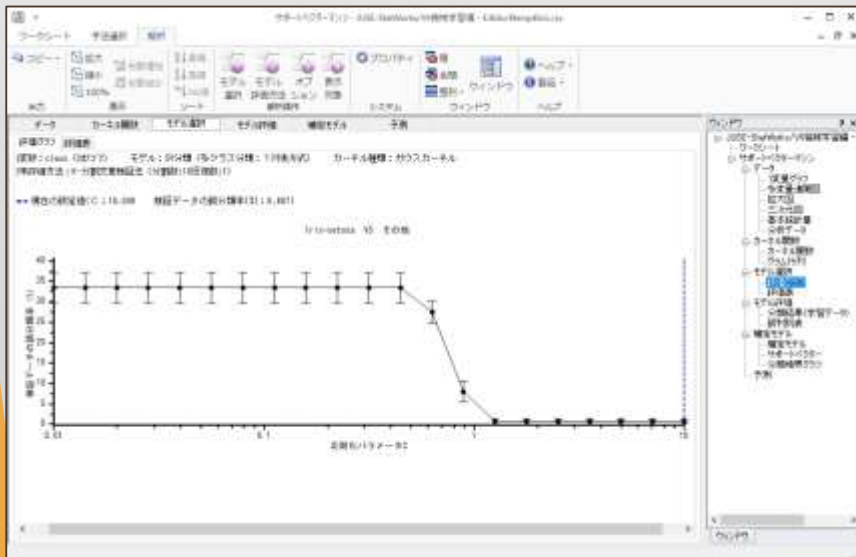
1. StatWorks/V5機械学習編のご紹介
2. デモンストレーション

StatWorks/V5機械学習編 のご紹介

JUSE-StatWorks/V5 機械学習編

スタンドアロン版
ネットワーク版

2019年4月25日発売
2019年6月27日発売



StatWorks/V5の製品構成

製品 手法群

総合編プレミアム

SEM因果分析編

因果分析

総合編

多変量解析

時系列解析

品質工学編

信頼性解析

品質工学

品質管理手法編

実験計画法

回帰分析

QC七つ道具編

QC七つ道具

新QC七つ道具

工程分析

機械学習編

機械学習

新規
追加



開発の背景

1. モノづくりの問題解決におけるデータの多様性の増加

- ▶ 情報システムの整備や工場等でのIoTの導入により、多様性に富んだデータを利用できる環境が整いつつある
- ▶ データによっては、従来のsqc手法ではうまく対処できなかったり、手間がかかる等の問題が生じることがある。その対処の一つが機械学習の活用。

2. 現場の技術者が使い易い機械学習ツールに対する強いニーズ

- ▶ TQMの理念は全員参加であり、現場の技術者が自分でデータ解析を行う
- ▶ 既存の機械学習ツールは専門家向けが多い
- ▶ StatWorks/V5に機械学習手法を搭載する強いニーズ

システムコンセプト

1. 現場の技術者のためのツール

- ▶ 現場の技術者が手軽に使える機械学習ツール
- ▶ マウスで全て操作可能
- ▶ 単体PC上で動作（サーバー・クラウド不要）

2. 全て自社開発

- ▶ 計算処理も自社開発
- ▶ ユーザー要望への対応可

3. StatWorksのコンセプトを引き継ぐ

- ▶ 豊富なグラフ機能
- ▶ 解析ストーリーに沿ったタブ・機能構成

機能・操作性の特徴

1. 既存製品と一体で使用可能

- ▶ 総合編等と機械学習編を同じPCにインストールするとメニューが統合される

2. 既存製品と同じ操作性

- ▶ ワークシートは共通
- ▶ 画面レイアウトも共通

3. 1,000変数(列) × 100,000サンプル(行) 以内のデータを解析可能

機械学習編
メニュー



搭載手法一覧



No	手法分類	グループ	解析手法
1	データ前処理	データクリーニング	データクリーニング
2		データ可視化	濃淡散布図, 密度プロット, 等高線図
3	教師なし学習	情報要約	カーネル主成分分析
4		層別	混合ガウス分布
5		外れ値検出	1クラスSVM
6		因果分析	glasso
7	教師あり学習	正則化回帰	リッジ回帰, lasso回帰, Elastic Net
8		分類・予測	サポートベクターマシン(SVM)
9			ランダム・フォレスト

搭載手法の特徴

1. モノづくりの問題解決で有効

- ▶ 従来のsqc手法ではうまく対処できなかった状況で特に有効
- ▶ 変数の数がサンプル数よりも多い状況（多重共線性が発生している状況）やデータの分布が正規分布から大きく乖離している状況などにも対応

2. 要因の検討でも使用可

- ▶ 変数選択が自動的に行われる，重要度が出力される など

3. 多変量解析の拡張として理解可能

搭載手法と多変量解析手法との対応

No	機械学習手法(A)	対応する多変量解析手法(B)	B→Aの主なメリット
1	カーネル主成分分析	主成分分析	複雑な構造のデータに対応
2	混合ガウス分布	非階層的クラスター分析(k-means)	所属確率の取得
3	リッジ回帰, lasso回帰, Elastic Net	重回帰分析・数量化Ⅰ類	変数の数>サンプル数, の場合(多重共線性) に対応
4	サポートベクターマシン(SVM)	判別分析・数量化Ⅱ類	汎化能力(新たなデータに対する予測能力)の向上
5	ランダム・フォレスト	多段層別分析(AID)	汎化能力の向上
6	1クラスSVM	MT法	正規分布に従わないデータに対応
7	glasso	グラフィカルモデルリング(GM)	多重共線性に対応

有効な活用場面

1. 中規模データの分析

- ▶ R, Pythonなどに比べ, 手軽に分析を実行可能
- ▶ 1,000変数×100,000サンプル以内のデータを分析でき, モノづくりの問題解決で使用する数値データの多くをカバー

2. 機械学習手法の教育

- ▶ 教育を行うための環境構築が容易
- ▶ マウスで操作できるので演習を自由に実施可能
- ▶ トヨタグループ様の機械学習セミナー, 日科技連の機械学習セミナー (2020年度開講) でも使用

3. ビッグデータの分析モデルの検討

- ▶ ビッグデータから抽出したデータを分析
- ▶ 分析の試行錯誤を行い易い

稼働環境

【スタンドアロン版・ネットワーク版クライアント】

製品	OS	ハードウェア
機械学習編	【64bit版】 Windows 10	【CPU】 Intel Core i5-4xxx以上 【メモリ】 8GB以上 【HDD】 1GB以上の空き容量 【ディスプレイ】 1024×768(XGA)以上 【ディスク装置】 DVD-ROMドライブ

【ネットワーク版サーバー】

製品	OS	ハードウェア
機械学習編 含む全製品	【64bit版】 Windows Server 2016 Windows Server 2012 R2 【32bit版, 64bit版】 Windows Server 2008 R2	【CPU】 1GHz以上(x86), 1.4GHz以上(x64) 【メモリ】 2GB以上(x86, x64) 【HDD】 約130MB以上の空き容量

製品価格

【スタンドアロン版】

製品	標準価格(税別)
機械学習編	240,000円
総合編 + 機械学習編セット	367,000円
総合編	168,000円

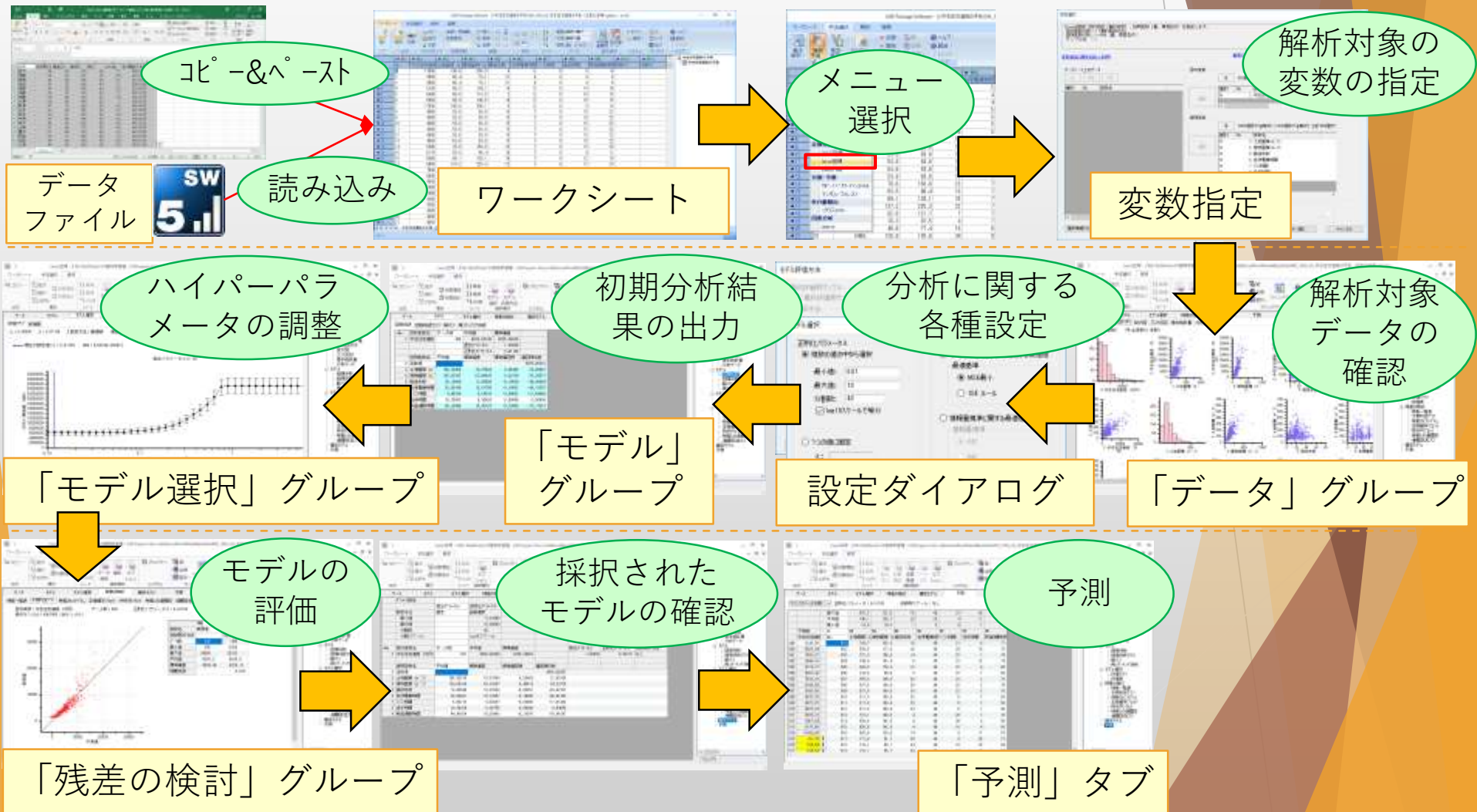
【ネットワーク版】

※同時起動数・インストール台数は5単位で増減

製品	同時起動数	インストール台数	標準価格(税別)
機械学習編	5	25	2,100,000円
総合編 + 機械学習編 セット	5	25	3,200,000円
総合編	5	25	1,450,000円

製品価格は弊社ホームページ上でご確認ください

基本的な分析手順(lasso回帰)



StatWorks/V5機械学習編 デモンストレーション

自動車のカタログスペックデータに対し、lasso回帰を用いて、目的変数「燃費」を他のスペック（17項目）から予測する式を構築する例をご紹介します。

機械学習に関する主催セミナー

▶今年度開講の機械学習セミナー

分類	セミナー名	開催期間	受講料 (税別)
操作入門 コース (半日間)	①正則化回帰 (lasso回帰, ElasticNet)	2019/7/24, 2019/10/29	9,000円
	②ランダム・フォレスト, サポートベクターマシン (SVM)	2019/8/22, 2019/11/21	9,000円
	③混合ガウス分布, glasso	2019/9/27, 2019/12/13	9,000円
手法マスターコース	問題解決のための機械学習入門	2020/1/15-16 (2日間)	70,000円

▶来年度も同様のセミナーを開講予定

おわりに

- ▶ 本発表では、StatWorks/V5機械学習編の特徴や機能をご紹介をしました。
- ▶ この後弊社 山田から、機械学習編の分類手法（サポートベクターマシン(SVM), ランダムフォレスト）を用いて、画像データを分析した例を紹介します。
- ▶ StatWorks/V5機械学習編が、皆様の機械学習の活用の一助となれば幸いです。

ご清聴どうもありがとうございました。

本著作物は原作者の許可を得て、株式会社日本科学技術研修所（以下弊社）が掲載しています。本著作物の著作権については、制作した原作者に帰属します。

原作者および弊社の許可なく営利・非営利・イントラネットを問わず、本著作物の複製・転用・販売等を禁止します。

所属および役職等は、公開当時のものです。

■公開資料ページ

弊社ウェブページで各種資料をご覧ください <http://www.i-juse.co.jp/statistics/jirei/>

■お問い合わせ先

(株)日科技研 数理事業部 パッケージサポート係 <http://www.i-juse.co.jp/statistics/support/contact.html>